Self-fulfilling prophecies in the classroom

Andrew J. Hill and Daniel B. Jones

Self-Fulfilling Prophecies in the Classroom

Andrew J. Hill
Montana State University

Daniel B. Jones
University of Pittsburgh

Do teachers' expectations directly impact student achievement? We draw on administrative data from North Carolina schools that report both student test scores and teachers' expectations of students' performance on these tests. Employing student fixed effects and instrumental variables strategies to overcome endogeneity concerns, we find that higher exogenously determined teacher expectations increase test scores for fourth to eighth graders. Impacts are suggestively larger for students in earlier grades and in self-contained classes with the same math and reading teacher.

I. Introduction

Do teachers' expectations of students impact student achievement? That is, are teachers' expectations self-fulfilling prophecies? These questions have been debated for decades. Rosenthal and Jacobson (1968) initiated much of the debate with their now classic experiment in which teachers were told some of their students had more academic potential, while, in reality, they were chosen at random by the researchers. Despite this, the randomly selected students displayed greater test score gains by the end of the year. This effect—teacher expectations causally impacting student achievement—has come to be known as the Pygmalion effect or Rosenthal effect. Other researchers have followed with numerous critiques and replication efforts, leading to mixed conclusions (e.g., Brophy 1983; Wineburg 1987). For instance, shortly after the publication of his first study, Rosenthal himself attempted to replicate his results at a different school in a different part of the country and found no evidence that students randomly assigned to the "high expectations" group performed any better (Evans and Rosenthal 1969). In a recent review of the literature, Jussim and Harber

Electronically Published July 12, 2021.

[*Journal of Human Capital*, 2021, vol. 15, no. 3]

(2005) note that, of the replications, "only slightly over one third demonstrated a statistically significant effect" (135).[1]

In this paper, we draw on a rich administrative data set and modern econometric techniques to shed new light on whether teachers' expectations are a self-fulfilling prophecy. Using student-teacher linked data from North Carolina, where we observe both how students perform on standardized state-written end-of-grade (EOG) tests and how their teacher expected them to perform, we combine an instrumental variables strategy with a rich set of fixed effects to estimate the causal impact of teachers' expectations on students' performance on the EOG tests.

Our approach hinges on constructing a measure of teachers' underlying optimism to use as an instrument for teachers' student-specific expectations. It is similar in spirit to the instrumental variable approach used to identify systematically more (or less) lenient judges in recent research on the impacts of judicial outcomes (e.g., Kling 2006; Dahl, Kostøl, and Mogstad 2014; Dobbie, Goldin, and Yang 2016). Our strategy is described in more detail in a later section, but, in short, we use student-level panel data and regress teachers' expectations on a variety of student characteristics, student fixed effects, and teacher fixed effects, and then take the teacher fixed effects as a measure of underlying teacher optimism. We omit data from year t when constructing a teacher's year t underlying optimism measure to ensure that it is not affected by the achievement of the teacher's current students but based only on past students. Intuitively, this measure captures the degree to which some teachers consistently over- or underestimate their students' achievement after conditioning on the types of students they teach. Using this as our instrument for teachers' student-specific expectations, our experiment essentially entails observing how a particular student's achievement varies as they are assigned to generally more or less optimistic teachers while progressing through elementary and middle school.

We find clear evidence of self-fulfilling prophecies in the classroom. The most robust specifications we report indicate that a teacher who has a one-level higher belief on a four-point scale about a student's mastery over the subject material increases that student's EOG score by roughly 0.07 standard deviations of a test score. Similar to Rosenthal and Jacobson's (1968) results, we find suggestive evidence that these effects are strongest in earlier grades. Also similar to their findings, we observe no evidence that the positive gains from higher expectations persist and generate gains in later academic years.[2]

[1] Some more recent studies aim to measure (rather than randomly assign) teachers' expectations in a natural setting; in these studies, researchers have argued that observed positive correlations between teacher expectations and student performance largely stem from accurate teacher expectations rather than self-fulfilling prophecies (e.g., Jussim 1989; Trouilloud et al. 2002).

[2] One exception to the lack of persistence, which we discuss in more detail in a later section of the paper, occurs when students are assigned to the same teacher 2 years in a row. Among

We conduct a number of tests aimed at probing the validity of our empirical approach. For instance, if our approach is truly capturing the causal impact of a teacher's expectations on test scores, then the expectations of a teacher that a student is assigned to in the future should have no influence on current test scores. We find that this is the case. This test and other tests are described in more detail in later sections.

Our paper primarily contributes to two strands of literature: first, a broad literature on teacher effectiveness and, second, a small literature in economics (and a larger literature outside of economics) on self-fulfilling prophecies. Regarding the former, there is now clear evidence on the importance of teachers and teacher quality in the education production function, both on student achievement (Rockoff 2004; Rivkin, Hanushek, and Kain 2005) and later-life outcomes (Chetty, Friedman, and Rockoff 2014). Less clear are the factors driving increased teacher effectiveness. A substantial body of literature has examined the role of observable qualifications. Clotfelter, Ladd, and Vigdor (2007, 2010) draw on rich models estimated in administrative data to document positive effects of teacher experience and credentials on student achievement. Other work suggests such characteristics cannot explain variation in teacher effectiveness. For example, while generally documenting the importance of teacher quality on student achievement, it has been shown that licensure test scores (Goldhaber 2007), National Board certification (Harris and Sass 2009), and education and credential type (Aaronson, Barrow, and Sander 2007) are less predictive of classroom performance. In estimating models that evaluate optimism as a fixed characteristic of teachers, our paper puts forth and provides evidence on the importance of a less immediately observable component of teacher effectiveness.

Regarding the literature on self-fulfilling prophecies, in an early paper focused on a single school district, Link and Ratledge (1979) draw on student test score data and a questionnaire that asked students their perception of their teacher's expectation of them. They find a positive correlation between test scores and students' perception of their teachers' expectations, but there is no attempt to account for potential endogeneity in their expectation measure. In a recent paper, Papageorge, Gershenson, and Kang (2020) analyze data from the Educational Longitudinal Study (ELS2002) where two teachers report their expectations about a high school student's likelihood of completing college. They leverage the disagreement between teachers to estimate the effect of expectations on college completion, ultimately finding that students who are expected to complete college by teachers are more likely to do so. Carlana (2019) shows that teachers with implicit gender biases widen stereotypical gender achievement gaps in their classrooms, providing another channel through which teacher beliefs affect student outcomes. Our results on the immediate

those students, the effect of positive expectations in the first year of a student-teacher match has an even larger effect on test scores in the second year.

impact of expectations on student achievement complement these findings. Outside of the context of education, another recent paper (Glover, Pallais, and Pariente 2017) provides evidence that ethnic minority grocery store employees exert less effort when quasi-randomly assigned to work for a manager who is biased against ethnic minorities (as measured by an Implicit Association Test), thereby reinforcing the biased managers' expectations.

The impact of teacher expectations on student achievement takes on renewed importance with recent evidence in economics that teachers have systematically lower expectations of ethnic minority groups, even when these lower expectations are inaccurate. Burgess and Greaves (2013) draw on administrative data from the United Kingdom where they observe both students' test scores and teachers' assessments of students' ability levels. Even with a rich set of controls (including students' prior test scores and poverty status), they find that teachers are significantly more likely to underestimate black students, reporting a subjective assessment that is lower than the students' actual performance in the relevant test. Gershenson, Holt, and Papageorge (2016) take advantage of the same feature of the ELS2002 mentioned above and, holding student characteristics constant (through student fixed effects), document that white teachers hold systematically lower expectations of black students than black teachers do; the same is not true of white students. Papageorge, Gershenson, and Kang (2020) find that teachers are overoptimistic about all students' likelihood of college completion, but the degree of excess optimism is larger for white students than for black students.[3] Van Ewijk (2011) conducted an experiment in Dutch schools where identical essays were graded by multiple teachers, but some teachers were randomly assigned essays with names suggesting that the student was an ethnic majority, while others were led to believe the essay was written by a minority. Despite the fact that the essays were identical, teachers reported lower expectations for the school track (university preparatory, vocational, etc.) the student would be able to attend when the name suggested the student was a minority. In a related study, Figlio (2005) argues that teachers use names as a signal of unobserved parental inputs and have lower expectations of children with names that sound like they were given by uneducated parents. More recently, Francis, de Oliveira, and Dimmitt (2019) perform an audit study in which they find that school counselors rate black females as being least prepared for an Advanced Placement course even with transcripts identical to their peers'.[4]

[3] These findings contrast with earlier findings suggesting that differences in teacher expectations of black and white students are eliminated after controlling for past performance and other observables (Ferguson 2003).

[4] Further evidence of teachers using characteristics other than academic performance to assess students is provided by Condron (2007), who finds that socioeconomic status and race affects reading group placement in kindergarten.

A closely related literature reveals evidence of teacher bias in grading, providing a different type of evidence that teachers' perceptions of student ability can be driven by factors other than accurate estimates of students' true underlying ability. Hanna and Linden (2012) conduct an experiment in India and find that teachers discriminate against students thought to be from lower castes, while Zavodny (2013) finds that teachers' assessments are negatively correlated with students' body weights after controlling for test scores. Some papers exploit differences in student scores under blind versus nonblind grading when grading can be subjective (e.g., an essay rather than a multiple choice exam). Both Lavy (2008) and Terrier (2016) document that male students' scores suffer with nonblind grading in comparison with blind grading, suggesting that teachers have lower expectations of male students. Mechtenberg (2009) provides a theoretical model documenting the potential for biased grading to generate self-fulfilling prophecies; the group subject to negative bias interprets lower grades as a signal of low ability or marginal returns to effort and exerts less effort in the future. Our paper is distinct from this literature in several ways. First, our goal is not to document bias in teachers' expectations. We focus on the impact of expectations on student test scores. Second, the test scores we take as an outcome variable are multiple-choice tests and are not graded by teachers, so there is no scope for teacher bias in grading. Instead, we study how teacher expectations impact student performance more generally when there is no opportunity for subjectivity in grading.

Differences in teacher expectations by race, ethnic group, or gender are not the primary focus of this paper, but these findings highlight the potential for (and, indeed, the presence of) variation in teacher expectations driven by factors other than simply variation in student ability. If it is true (as the above papers suggest) that teachers have lower expectations of minority students, then understanding how these (often-mistaken) teacher expectations can directly impact student performance is of substantial policy importance, especially given the persistence of the black-white achievement gap in the United States (Hedges and Nowell 1998). Our goal in this paper is to provide clear evidence of the role of teacher expectations using new comprehensive administrative data and empirical techniques.

Finally, it is worth looking to the literature to understand, theoretically: How might teacher expectations influence student achievement? A basic theoretical framework is described by Jussim, Robustelli, and Cain (2009)—and reviewed in work by Gentrup et al. 2020—in which, first, a teacher forms inaccurate expectations; second, the teacher's expectations impact the teacher's behavior toward the students; and third, as a result of differential treatment toward high- and low-expectation students, students respond in a way that ultimately confirms the inaccurate expectations of the teacher. With regards to the first piece, recent evidence from Rangel and Shi (2020) sheds light on one way in which teachers may form inaccurate expectations; namely, they document that beginning teachers systematically overevaluate the abilities of white students and underevaluate the ability of black students, consistent with other findings noted above. Thus, one piece of the inaccurate expectations stems from biased views toward particular groups on the part of the teacher. However, they then go on to document the malleability of teachers' expectations; the gap in teachers relative expectations of black and white students can either be exacerbated or diminished in later years based on the actual relative performance of black and white students in the teachers' first years of teaching. Next, Rosenthal (1973) highlights four ways in which teachers may change their behavior based on their expectations: for high-expectation students, (1) teachers may provide a more supportive climate (climate channel), (2) teachers may provide feedback directly related to performance rather than, for instance, behavior (feedback channel), (3) teachers may direct more attention toward or offer more challenging material to the students (input channel), and (4) teachers may provide more opportunities for the students to succeed (output channel). Notably, all four channels could occur at a class-wide rather than student-by-student level, which is relevant, as our instrumental variable analysis implicitly tests the impacts of being assigned to a teacher that holds higher expectations generally. A meta-analysis by Harris and Rosenthal (1985) revealed that all four channels are correlated with teacher expectations as well as student achievement, but climate and input revealed the strongest relationships.[5]

II. Data

We use data on the population of third-to-eighth-grade students and teachers in public North Carolina elementary and middle schools between 2007 and 2013 obtained from the North Carolina Education Research Data Center (NCERDC). The data include rich student and teacher demographics, as well as school and district characteristics. The unique feature of the data that we exploit in this paper is that—in every year and for every student—we observe both the student's EOG test scores and the level of achievement the teacher anticipates the student will attain in the tested subject.

Students in grades 3–8 take two EOG tests each year: math and reading.[6] The EOG tests are multiple-choice exams and are standardized across the state. This removes any opportunity for teachers to manipulate either the difficulty or the grading of the tests. Our data are organized as a panel with observations at the student-by-year-by-subject (math and reading) level.

We provide a detailed description of the variables for teacher expectations in the data given their uncommon use. For both math and reading

[5] Recent work from Gentrup et al. (2020) document that inaccurately high (or low) teacher expectations are associated with differences in the feedback provided to students, with higher-expectation students receiving more performance-related feedback. However, in their setting, they cannot confirm that this channel mediates the relationship between teacher expectations and student achievement.

[6] An EOG test on science is administered in fifth grade. Because it is not administered each year, we do not include results from the science test in our analysis.

(even if the same teacher teaches both subjects, as is often the case in elementary and middle school), the teacher judges the student's ability to be one of four levels: insufficient mastery of knowledge and skills to be successful at the next grade level (achievement level I); inconsistent mastery, minimally sufficient for the next grade level (level II); consistent mastery and well prepared for the next grade level (level III); and superior mastery, "clearly beyond that required to be proficient at grade-level work" (level IV). This expectation is reported by the teacher at the end of the academic year just after the student takes the EOG test but before the test scores are made available to teachers. It is important to note that in instructions to teachers, they are told to base their judgment "solely on mastery" of the subject matter and not on grades received throughout the year as "grades are often influenced by factors other than pure achievement" (https://casetext.com/regulation/north-carolina-administrative-code/title-16-state-board-of-education).[7]

It is worth briefly discussing the fact that the teacher expectation is reported at the end of the year. This is one sense in which our data is different than that available from classic experiments (e.g., Rosenthal and Jacobson 1968), where teachers' expectations are manipulated at the beginning of a school year. Thus, in practice, we are testing whether having higher expectations throughout the year impacts a student's final achievement, implicitly assuming that the belief reported by teachers at the end of the year is a reasonable (if relatively well-informed) proxy for the expectation the teacher formed throughout the year. Recall, in addition, that we omit data from year t when constructing year t underlying optimism, so we are excluding potential feedback from current student performance into the measure of current teacher expectations; to be clear, the measure is based on teachers' assessments of students in years other than t. In further analysis, we take advantage of the large sample size available in our data and estimate the impact of expectations reported at the end of year t on students' achievement in year $t + 1$ for the small subset of students who have the same teacher in consecutive years. That exercise, which is arguably closer to the classic experiments on the question, yields results similar to our main results.

Descriptive statistics of the primary variables are reported in table 1. Column 1 describes the teachers, and column 2 describes the students

[7] The full instructions to teachers are as follows: "The [subject (math or reading)] teacher is to identify each student who, in the [subject] teacher's professional opinion, clearly and consistently exemplifies one of the achievement levels listed. If a student is not a clear example of one of the listed achievement levels, circle 9 in col. E is to be coded. The [subject] teacher should base this response for each student solely on mastery of [subject material]. The [subject] teacher may elect to use grades as a starting point in making these assignments. However, grades are often influenced by factors other than pure achievement, such as failure to turn in homework. The [subject] teacher's challenge is to provide information that reflects only the achievement of each student in the subject matter tested. The [subject] teacher should therefore rely chiefly on professional experience about what is the appropriate achievement level" (http://wcstesting.pbworks.com/f/1011+Gr+3-8+EOG+TAM.pdf).

TABLE 1
DESCRIPTIVE STATISTICS—MEAN (Standard Deviation)

	Teachers (1)	Student-Subject-Years (2)
Female	.88	.50
Ethnicity:		
White	.85	.59
Black	.13	.23
Hispanic	.00	.11
Asian	.01	.02
Other	.01	.05
Teacher-specific variables:		
Value added (z-score):		
90th–10th percentile	2.17	
75th–25th percentile	1.07	
Graduate degree	.31	
	(.46)	
Experience (years; teacher's max over sample period)	13.84	
	(8.99)	
Student-specific variables:		
End-of-grade test score (z-score)		
90th–10th percentile		2.55
75th–25th percentile		1.33
Anticipated achievement level:		
I (insufficient mastery)		.04
		(.20)
II (inconsistent mastery)		.19
		(.39)
III (consistent mastery)		.47
		(.50)
IV (superior mastery)		.29
		(.45)
Scale from I to IV		3.01
		(.81)
Same math and reading teacher		.57
		(.50)
Observations	16,698	2,547,174

Note.—"Same math and reading teacher" is an indicator with a value of one when the student has the same math and reading teacher in a given grade and zero otherwise.

(all elementary and middle school students in public schools in North Carolina during the study period).

The vast majority (88%) of the teachers in the sample are female, which is expected, given our focus on elementary and middle schools. Almost five out of six teachers are white, almost one in six teachers are black, and other ethnicities constitute only 2% of the teachers in our sample. Taken together, white females are vastly overrepresented in the teacher labor market in North Carolina. We construct our measure of teacher value added in a standard way by regressing student test scores on teacher fixed effects and a vector of other controls as detailed by Hill and Jones (2020), but leaving out year t when estimating value added for a given teacher in year t. The estimated teacher fixed effects are then standard normalized across the universe of teachers in North Carolina and serve as our measure of teacher quality. We omit year t when estimating the value added for a given teacher in year t because when analyzing the teacher's expectations of student i and student i's achievement, we do not want that student to contribute to the estimation of the teacher value-added measure we use as a control. Thus, although the measure is not in fact time invariant, it should be thought of as capturing a teacher's time-invariant ability to generate test score gains. Given that the variable is a z-score by construction, we report statistics other than the mean and standard deviation: the difference between the 90th and 10th percentiles is about 2 standard deviations, and the difference between the 75th and 25th percentiles is about 1 standard deviation. (Though not reported in table 1, we also construct a year-specific value-added measure, which we lag and include as a control for recent teacher ability.) One-third of teachers have some form of graduate degree, while the mean years of experience (assigning each teacher their maximum years of experience) is 13.8.

Turning to column 2, students are approximately 50% female, 59% white, 23% black, 11% Hispanic, 2% Asian, and 5% other ethnicity. Similar to the measures of teacher value added, the EOG test scores are standard normalized across the universe of North Carolina students . Student test z-scores are expectedly more dispersed than teacher value-added z-scores: the difference between the 90th and 10th percentile is 2.6 standard deviations, and the difference between the 75th and 25th percentile is 1.3 standard deviations. The variable of most interest for us is teachers' anticipated level of achievement. We see that teachers expect 4% of students to show insufficient mastery of the skills and knowledge in a given subject (achievement level I); 19% of students to show inconsistent mastery (level II); 48% students to show consistent mastery (level III); and 29% to show superior mastery (level IV). The next row shows that the mean and standard deviation of the four-point scale are 3.0 and 0.8, respectively, so a one-level change in beliefs corresponds to about a 1.25 standard deviation change. Finally, about 57% percent of student-subject-year observations involve students who have the same math and reading teacher in a given year, a proxy for being in a self-contained class rather than having subject-specific teachers.

Table A1 (tables A1–A6 are available online) provides descriptive statistics on students and teachers, split into two samples: teachers above and below the median of our measure of underlying teacher optimism (panel A) and the students of these teachers (panel B). As will be explained when we describe the construction of the teacher optimism measure in the next section, teacher and student characteristics are factored out when constructing our optimism measure to ensure that it is not capturing something else about teachers or students. Any differences in the mean demographics of students taught by less or more optimistic teachers are expectedly very small in magnitude, although very precisely estimated. We do, however, observe generally higher test scores for students taught by more optimistic teachers—the relationship of interest more fully explored in the subsequent regression analysis—and confirm that more optimistic teachers have higher expectations of their students (after factoring out all of their characteristics). In addition, we find that 69% of students in the sample are taught by both a teacher with above-median optimism and a teacher with below-median optimism, confirming that students are typically exposed to considerable variation in teacher optimism.

We now turn to explaining how we use these variables to identify the causal effect of teacher expectations on student achievement.

III. Empirical Methodology

Before discussing our empirical specifications, we begin by describing the general aims of our empirical strategy. Consider the following education production function describing the educational output of student i in subject s (math or reading) in year t, where student i is assigned to teacher j:

$$\begin{aligned} \text{TestScore}_{ijst} = {} & \alpha \text{StudentSpecificTeacherExpectations}_{ijst} \\ & + \beta \text{TestScore}_{is(t-1)} + \gamma \text{StudentCharacteristics}_i \\ & + \delta \text{TeacherCharacteristics}_j + \eta \text{TeacherQuality}_{jt} \\ & + \theta \text{PeerCharacteristics}_{ijt} + e_{ijst}. \end{aligned} \tag{1}$$

The function captures several established inputs affecting standardized EOG test scores: previous test scores, fixed student and teacher characteristics (such as race and gender), teacher quality, and peer characteristics. Most importantly, it also includes the impact of teacher expectations, our explanatory variable of interest. Note, again, that—like the test score—teacher expectations are student and subject specific. For our purposes, a teacher's expectation of their student is defined as the level of mastery (on a I–IV scale) that student i's teacher expects that the student's performance on the EOG test will display (as discussed in sec. II).

The objective of this paper is to estimate α, the causal effect of year-, student-, and subject-specific teacher expectations on test scores. The econometric challenge is clear; we need to isolate variation in student-specific teacher expectations that is uncorrelated with any component of the error term e_{ijt} (or, in other words, uncorrelated with any input that is not otherwise accounted for in the test score model).

Our approach to dealing with this challenge is twofold. First, in our main empirical specifications, we take advantage of the panel structure of the data and include student-by-subject fixed effects or lagged test scores.[8] This eliminates the influence of time-invariant student characteristics that

[8] We also report a specification where we include both student fixed effects and lagged test scores, although models with both lagged dependent variables and unit fixed effects are known to have undesirable properties, so we do not include both controls together in the main specification.

could impact both teacher expectations and students' test scores (e.g., fixed student ability or motivation in the classroom) and also deals with potential nonrandom sorting into classrooms that may be a function of fixed student characteristics (e.g., the student's family; Qureshi and Ost 2018) or past student performance (tracking). The second concern is the presence of shocks in period t (unobservable to the econometrician) that affect both a student's test score and their teacher's expectation of their test score. For example, a student who experiences a negative health shock and misses a considerable amount of school is likely to obtain a lower EOG test score given the missed material. At the same time, the student's absence will be known to the teacher, so the teacher is likely to expect that student to have a lower EOG test score. It would clearly be a mistake to attribute the student's lower test score to the teacher's lower expectation of their test score in this circumstance. We implement an instrumental variables strategy to deal with threats to identification of this type, which we now describe in more detail.

A. *Construction of Our Instrumental Variable*

Broadly, our approach is to construct a measure of fixed underlying teacher optimism (or pessimism) about students' test scores—controlling for the types of students a teacher faces—and use this as an instrument for student-specific teacher expectations in the test score model (an empirical model similar to eq. [1]). Our experiment entails exploiting within-student variation in teacher optimism as the student progresses through grades 3–8. The shocks to student-specific teacher expectations are driven by assignment to teachers across grades that are more (or less) optimistic in general.

To construct the main version of our instrument, we regress student-specific teacher expectations on teacher fixed effects and a variety of student-, cohort-, and (time-varying) teacher-level controls that could also impact teachers' expectations. The resulting estimates of the conditional teacher fixed effects are our measure of underlying teacher optimism. Note that this is not an altogether new approach to measuring a teacher characteristic: the construction of our fixed teacher optimism measure is mechanically similar to the estimation of teacher-level value added. As in our estimation of value added, in order to construct a measure of a teacher's optimism that might impact students in year t, we exclude data from year t. Again, we aim to avoid the possibility that student i (or that student's peers) contribute to our estimate of a teacher's general level of optimism, while also testing whether that student is in turn impacted by the constructed optimism measure.[9] Our instrument also bears some similarity to instruments constructed by Kling (2006) and Dahl,

[9] In practice, our results are very similar without making this leave-year-out restriction. Note also than an alternative approach would be to use only years up until $t - 1$ to construct the year t optimism measure. Results are similar using this approach, although we do not do

Kostøl, and Mogstad (2014), who use past decisions from a judge as an instrument for a judge's leniency.

Specifically, taking teacher j's expectation of how student i will perform on the subject s test in year t as the outcome variable, we estimate

$$\begin{aligned}\text{StudentSpecificTeacherExpectations}_{isjt} = {} & \kappa f(\text{TestScore}_{is(t-1)}) \\ & + \pi\text{StudentRace} \times \text{TeacherRace}_{ijt} \\ & + \psi_1\text{TimeVariantTeacherQuality}_{j(t-1)} \\ & + \psi_2\text{GeneralTeacherQuality}_{jt} \\ & + \sigma\text{PeerCharacteristics}_{ijt} + u_j + \rho_{is} \\ & + [\text{Subj}_s \times \text{Year}_t \times \text{Grade}_g\text{FEs}] + v_{isjt}.\end{aligned} \tag{2}$$

As discussed above, StudentSpecificTeacherExpectations$_{isjt}$ is the teacher's reported beliefs about a student's level of mastery over a subject (reported on a I–IV scale). The model includes a variety of controls to ensure that the estimated teacher fixed effects are stripped of any factors affecting both student achievement and student-specific teacher expectations other than fixed, underlying teacher optimism (or pessimism).

First, student-by-subject fixed effects ρ_{is} account for any time-invariant student characteristics that could conceivably impact teachers' expectations (whether observed in the data or not). For example, existing literature suggests that teachers have higher expectations of white male students (van den Bergh et al. 2010; Harber et al. 2012; Burgess and Greaves 2013; Gershenson, Holt, and Papageorge 2016). In addition, unobservable student characteristics such as motivation and parental involvement are also likely to affect both student achievement and teacher expectations. All of these will be captured in the student-by-subject fixed effects. Including student-by-subject fixed effects (rather than simply student fixed effects) allows for the possibility that teacher expectations for a particular student (or type of student) can vary by subject. Terrier (2016), for instance, finds that teachers have biased expectations of female students in math only.

Second, we include flexible controls for students' performance on the same-subject EOG test for the previous grade TestScore$_{is(t-1)}$, which is our best proxy for students' (time-varying) level of proficiency and preparation for their current EOG test. This enters as a cubic function, although results are very similar if it enters linearly or nonparametrically. In addition, the student-by-subject fixed effects and lagged test scores jointly account for many forms of potential nonrandom sorting into classrooms that could bias estimates if correlated with teacher expectations.

Third, we interact student race with teacher race (StudentRace $\times$ TeacherRace$_{ijt}$). This is motivated by an emerging literature that finds that the interaction between student and teacher race affects teacher

---

use this as our primary optimism measure because optimism measures for earlier years in the sample with few preceding years are less precisely estimated.

expectations, in particular, that white teachers have lower expectations of black students (Gershenson, Holt, and Papageorge 2016). Note that student and teacher race do not enter directly as they are subsumed by corresponding student and teacher fixed effects.

Fourth, teacher quality may affect student-specific teacher expectations. If high-quality teachers know they are above average, they may expect their students to perform better in comparison with how they would have performed with an average teacher. We therefore include a value-added measure of teacher quality spanning a teacher's entire career in North Carolina (GeneralTeacherQuality$_{jt}$, which—as noted—excludes students from the current year in being estimated). Because teacher quality can evolve over time, we also include the lagged value of a year-specific value-added measure (TimeVariantTeacherQuality$_{j(t-1)}$).

Fifth, peer characteristics may affect teachers' expectations of their students. For example, a teacher may believe that a generally high-ability class generates positive spillovers for all students in the class, or alternatively, a teacher's expectation of a given student may depend on their rank in the class. There is emerging evidence in other educational contexts that assessments relative to one's peers matter (Kinsler and Pavan 2016). To account for this, we control for the class-level average (leaving out student i) of race composition, gender composition, and previous-year EOG scores (captured in PeerCharacteristics$_{ijt}$).

We also include the full interaction of subject, year, and grade fixed effects to account for any year- or grade-specific variation in student cohort quality or the difficulty of the EOG test.

This leaves us with the teacher fixed effects u_j. Having differenced out the various expectation-influencing covariates described above, higher general expectations (as indicated by a higher value of the teacher fixed effect) from one teacher cannot simply be explained by teaching better students (as this is captured by lagged student scores and student fixed effects), demographic match with students (captured by student-teacher race interactions), increased belief in ones' own ability to improve student scores (captured by the teacher quality measures), or anything else captured by the controls listed above. We therefore consider the teacher fixed effects to capture teachers' underlying (and exogenously determined) tendency to over- or underestimate students throughout their career: a measure of teacher optimism. After estimating equation (2) separately for each year (leaving out the current year as we do so), we standard normalize the estimated teacher fixed effects to be mean 0 and standard deviation 1 in the population of third-to-eighth-grade teachers and use them as an instrument for student-specific expectations in our analysis.

B. Initial Probing of the Validity of Our Constructed Instrument

Here, and in later sections, we subject the teacher optimism measure to scrutiny to assess its validity as a measure of teacher optimism and an instrument for teachers' expectations. Our measure would fail if there is any systematic correlation between two teachers' expectations of the same student. This would reveal that the teacher optimism measure is still influenced by individual student characteristics and does not capture some exogenously determined tendency for a teacher to expect their students to perform better (or worse). To probe these concerns, we report the correlation between the optimism measure of student *i*'s teacher in year t (and grade g) and the optimism measure of the same student's teacher in year $t + 1$ (and grade $g + 1$). We do so for both the main version of our teacher optimism measure (described in the previous subsection) and a simpler version, where the estimation of the optimism measure controls only for subject-grade-year fixed effects (but no student or teacher characteristics beyond teacher fixed effects). Our goal is to show that our main measure succeeds where the simpler measure fails.

Indeed, this is what we find in the correlation matrix reported in table 2. For the simple measure, there is a relatively strong correlation between the general optimism of student *i*'s teacher in year t ("teacher optimism simple version") and the optimism of the same student's teacher the next year ("lead(teacher optimism simple version)"). In fact, the strength of the correlation (0.44) is very similar to the strength of the correlation between the simple version of the measure and our main version of the measure (0.50). A teacher fixed effect estimated without differencing out the full range of covariates employed in the main version of the measure fails this test. There are a variety of explanations for this, such as the simplest version of the teacher optimism measure carrying with it some information about the general level of optimism of all teachers in student *i*'s school.

Our main version of this measure, on the other hand, does not fail this test. Turning attention to the correlation between "teacher optimism main version" and "lead(teacher optimism main version)," there is essentially no relationship (correlation coefficient $= -0.06$). That is, being assigned to an "optimistic" teacher in grade g has no relationship with the level of optimism of the same student's teacher in grade $g + 1$.[10]

We provide an instrument balance test in table 3. We report results from regressing the simple and main teacher optimism measures on teacher characteristics at the teacher-subject-year level. Showing that our measure of teacher optimism is uncorrelated with observable teacher characteristics bolsters the argument that it is uncorrelated with unobservable teacher

[10] In a related test, we also find that the expectations of math and reading teachers for a given student in a given grade are moderately correlated (0.5) when using the simple version of teacher expectations but considerably less correlated (0.2) when using the main version of teacher expectations. This test is weaker than the one reported above because it is only applicable when students have different teachers for math and reading in the same grade, which considerably reduces the sample, and because same-grade math and reading teachers are likely to communicate more frequently than different-grade teachers about shared students.

TABLE 2
Student-Level Correlations between Current and Future Teacher Optimism Measures

	Teacher Optimism		Lead(Teacher Optimism)	
	Simple Version	Main Version	Simple Version	Main Version
Teacher optimism:				
Simple version	1			
Main version	.50	1		
Lead(teacher optimism):				
Simple version	.44	−.02	1	
Main version	.01	−.06	.50	1

Note.—"Teacher optimism main version" is our main optimism measure, estimated according to eq. (2). "Teacher optimism simple version" is estimated in a similar way to that denoted in eq. (2) but with no controls other than subject-grade-year fixed effects. Data are at the student-by-subject-by-year level. "Lead(teacher optimism)" therefore measures the optimism of the teacher that student i faces in year $t + 1$, while "teacher optimism" measures the optimism of student i's current (year t) teacher.

characteristics. Column 1 considers the simple version of the instrument (as described above). It is clear that this naïve teacher optimism measure is not balanced; the p-value on the F-test of significance of the parameters is <.001. On the other hand, column 2 reveals that our main measure of teacher optimism is balanced across teacher observables; it is predicted by none of the teacher characteristics, and we cannot reject the null hypothesis that all estimated coefficients are simultaneously equal to zero.

It is important to emphasize that the simple tests reported in tables 2 and 3 do not confirm that our measure is valid; it is not possible to do so definitively. Nonetheless, it is important to document that our measure passes these tests to increase confidence in its validity.

C. *Empirical Specifications Employed in the Main Analysis*

Having described the construction of our instrument, we now briefly describe the empirical specifications employed in the next section of the paper. Our empirical specifications are similar to the education production function presented in equation (1). In particular, in our richest specification, we estimate the following as the second stage of a two-stage least squares (2SLS) approach:

$$
\begin{aligned}
\text{TestScore}_{ijst} = {} & \alpha\,\widehat{\text{Expectations}}_{ijst} + \psi_1 \text{TimeVariantTeacherQuality}_{j(t-1)} \\
& + \psi_2 \text{GeneralTeacherQuality}_{jt} + \boldsymbol{\delta}\textbf{TeacherCharacteristics}_j \\
& + \pi \text{StudentRace} \times \text{TeacherRace}_{ijt} + \boldsymbol{\theta}\textbf{PeerCharacteristics}_{ijt} \\
& + [\text{Subj}_s \times \text{Year}_t \times \text{Grade}_g \text{FEs}] + \rho_{is} + e_{ijst}.
\end{aligned} \tag{3}
$$

TABLE 3
Instrument Balance Test: Correlation between Teacher Optimism Measures and Teacher Characteristics

	Teacher Optimism Simple Version (1)	Teacher Optimism Main Version (2)
Teacher characteristics:		
Female	−.002	−.039
	(.021)	(.028)
Black	−.219***	−.034
	(.024)	(.031)
Other ethnicity	−.091*	.013
	(.051)	(.074)
Value added	.402***	−.009
	(.011)	(.011)
Graduate degree	.023*	.002
	(.014)	(.018)
Experience (years)	.003***	−.001
	(.001)	(.001)
Observations (teacher-subject-year)	58,720	58,720
R^2	.450	.138
F-test of joint significance:		
p-value	.000	.512

Note.—Columns 1 and 2 consider different teacher optimism measures. "Teacher optimism main version" is our main optimism measure, estimated according to eq. (2). "Teacher optimism simple version" is estimated in a similar way to that denoted in eq. (2) but with no controls other than subject-grade-year fixed effects. All specifications include subject-grade-year and school fixed effects. Robust standard errors (clustered at the teacher level) are in parentheses.
* $p < .10$.
*** $p < .01$.

The outcome variable TestScore_{ijst} is student i's score on the subject s EOG in year t, and $\widehat{\text{Expectations}}_{ijst}$ is the predicted value of student-subject-year-specific teacher expectations from a first-stage where we use our constructed teacher optimism measure (the standard-normalized teacher fixed effects estimated from eq. [2]) as an instrument for the reported expectations for the student in question.[11]

Both the first and second stages include student-by-subject fixed effects (ρ_{is}). As a result, we are leveraging within-student variation in teacher expectations and test scores. Student-by-subject fixed effects (rather than simply student fixed effects) allow for the possibility that unobserved student characteristics affect math and reading achievement differently, which would be true if a student had a taste for math but a distaste for reading, for example. Other controls in the first and second stages parallel

[11] To clarify, eq. (2) is not the first-stage of the 2SLS estimator. Instead, eq. (2) provides the estimates of teachers' fixed tendency to over- or underestimate students, through the estimated teacher fixed effects in that equation. We then use these as the instrumental variable in eq. (3). The first stage in the 2SLS estimation procedure includes all the same controls as the second stage but takes actual expectations for a particular student on the left-hand side and the teacher optimism measure on the right-hand side.

those included in our construction of the teacher optimism measure: a vector of teacher characteristics (race, gender, graduate degree indicator, score on teacher qualification tests), two teacher quality measures (value added from the previous academic year and a measure capturing teacher quality across a teacher's career), a vector of peer characteristics (ethnic composition of classmates, gender composition of classmates, average lagged EOG score of classmates), teacher-student race interactions, the full interaction of subject, year, and grade fixed effects, and, in some specifications, lagged test scores.[12] Standard errors are clustered at the teacher level as that is the effective level of treatment.

An additional consideration stems from some teachers not teaching self-contained classes but rather teaching specific subjects to different classes. Teacher expectations may have different impacts on test scores when a student has the same teacher for most of the day in comparison with when they have different teachers for different subjects. This is most likely to be a factor for students in the higher grades in our sample: sixth to eighth graders. We cannot observe student and teacher assignment for all subjects, so we proxy for being in a self-contained class with an indicator for having the same math and reading teacher. This indicator is included in an extended model both linearly and interacted with our measure of teacher expectations to capture this form of heterogeneity. It becomes particularly important to include when considering whether there is heterogeneity in the effect of teacher expectations by grade level; self-contained classes become less likely as a student progresses to higher grades, so without controlling for this, we could conflate a self-contained class effect with a grade effect.

IV. Results

We begin by presenting ordinary least squares (OLS) estimates (in some cases, with student fixed effects) of the relationship between teachers' student-specific expectations and students' test scores in panel A of table 4. Column 1 reports the results of a very simple specification with no controls other than teachers' expectations and subject-by-year-by-grade fixed effects (which are included in all six specifications in panel A). We expect the estimated coefficient resulting from that specification to be substantially upwardly biased, as we make no attempt to eliminate the influence of factors that lead to high test scores and high teacher expectations (e.g., a generally highly motivated student, or—in this case, where we have not even controlled for lagged test scores—simply the fact that the teacher has at least some approximate sense of the student's ability level). Indeed, the coefficient suggests that raising teacher expectations of a student by one level (about 1.25 standard deviations) increases the student's test

[12] Lavy and Schlosser (2011) document the impact of cohort gender composition on achievement. Hill (2017) provides similar evidence at the college level. Burke and Sass (2013) document the impact of classmates' average ability level on achievement.

TABLE 4
Estimating the Impact of Teacher Expectations on Achievement

	(1)	(2)	(3)	(4)	(5)	(6)
	A. OLS					
Expectations	.837***	.281***	.076***	.090***	.074***	.074***
	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)
	B. 2SLS, First Stage					
Teacher optimism	.101***	.096***	.097***	.097***	.097***	.098***
	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)
	C. 2SLS, Second Stage					
Expectations	.086***	.019	.075***	.086***	.069***	.067***
	(.031)	(.012)	(.009)	(.009)	(.006)	(.006)
Hausman, OLS vs. 2SLS:						
Test statistic	586.44	489.42	.01	.88	15.22	2.33
p-value	.00	.00	.92	.92	.22	.99
	D. 2SLS, Second Stage Plus Additional Interaction					
Expectations	.205***	.039**	.057***	.068***	.056***	.055***
	(.040)	(.017)	(.014)	(.014)	(.009)	(.009)
Expectations × same math and reading teacher	−.212***	−.035**	.033**	.032**	.023*	.022*
	(.045)	(.018)	(.016)	(.016)	(.012)	(.012)
Lagged EOG (cubic)		X		X		
Student × subject fixed effects			X	X	X	X
Teacher characteristics					X	X
Cohort characteristics						X
Observations	2,547,174	2,547,174	2,547,174	2,547,174	2,547,174	2,547,174

Note.—The dependent variable in panels A, C, and D is end-of-grade (EOG) test score (*z*-score); in panel B, it is expectations. The Hausman test in panel C compares the ordinary least squares (OLS) estimates in panel A and the two-stage least squares (2SLS) estimates in panel C; the null hypothesis is that the OLS and 2SLS estimates are different. "Same math and reading teacher" in panel D is an indicator that is one when the student has the same math and reading teacher in a given grade and zero otherwise. Panel D includes an additional interaction with same math and reading teacher indicator. Robust standard errors (clustered at the teacher level) are in parentheses.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

score by nearly an entire standard deviation. Simply controlling for a student's lagged score (as in col. 2) cuts this coefficient down to less than one-third of a standard deviation. The coefficient shrinks even further when student-by-subject fixed effects are included both without lagged test scores (col. 3, a preferred specification) and with lagged test scores (col. 4, although recall that models of this structure have undesirable properties) but does not change much more as additional controls are added (cols. 5 and 6).

We now turn to the 2SLS estimation. Panel B reports the first stage from our 2SLS approach: the relationship between the main version of our constructed teacher optimism measure and the teachers' reported student-specific expectations. In all six specifications (across which we vary the same sets of additional controls as in panel A), the coefficient on our instrument is positive and very precisely estimated. A 1 standard deviation increase in the constructed teacher optimism measure increases the reported expectation measure by roughly one-tenth of a level on the I–IV scale of expected mastery. In other words, as should be expected given that teachers will have reasonably accurate expectations of students, many students are essentially unaffected by being assigned to a more optimistic teacher. However, some share of students who may otherwise be close to a cutoff between levels of mastery (level I vs. level II, etc.) will be perceived as a higher level by more optimistic teachers. In turn, our instrumental variables approach tests whether those students experience achievement gains as a result. It is also worth noting that the magnitude of the coefficient in the first stage is essentially unaffected by how many (or how few) controls are included, providing further suggestive evidence that our constructed instrument captures an exogenously determined measure of teacher optimism.

Panel C reports our main results: the second-stage estimates of our 2SLS estimation. In these results, we report the causal impact of higher teacher expectations (which have been instrumented for using our constructed overall teacher optimism measure) on students' EOG scores. Columns 1–6 gradually introduce controls in the same manner as panels A and B. Column 1 is the simplest specification, which includes only the (instrumented) teacher expectation and subject-by-year-by-grade fixed effects on the right-hand side. Although our preferred estimates include student-by-subject fixed effects and additional controls (which we will turn to momentarily), it is worth noting that the coefficients in columns 1 and 2 of panel C are much smaller than their OLS counterparts (cols. 1 and 2, panel A). This confirms that the relationship between teacher expectations and scores would suffer from an upward bias absent any attempt to deal with endogeneity.

As we move to specifications in columns 3–6, which include student-by-subject fixed effects, we observe a positive and statistically significant effect of teacher expectations that are somewhat similar in magnitude to the parallel OLS specifications. (Note that the statistical similarity of the OLS and 2SLS results in these columns are confirmed by Hausman tests comparing the estimates reported below.) This suggests that the upward bias in the OLS specifications evident in columns 1 and 2 is largely eliminated by the inclusion of student-by-subject fixed effects. Still, given the potential for unobservable year-specific shocks to both student achievement and teacher expectations, the 2SLS estimates are preferred. Ultimately, our richest—and main—specification is reported in column 6; there, we control for student-by-subject fixed effects, teacher characteristics, student-teacher race interactions, and cohort characteristics. The clear, positive impact of teacher expectations is seen even in this model.

We estimate that increasing teachers' expectations by one level of mastery (being perceived as a student who can master the material rather than simply attain inconsistent mastery or being perceived as a student who can attain superior mastery rather than simply mastery) increases student achievement by 0.067 standard deviations of a test score. In short, teacher expectations have a causal positive effect on student achievement, confirming the Rosenthal effect.[13]

The effect is moderate in size relative to other factors affecting elementary and middle school test scores. A 1 standard deviation increase in teacher expectations (0.8 of one level on the four-point scale) increases test scores by approximately 0.05 standard deviations, which is about equivalent to the test score gain from being in a repeat match with a teacher between third and fifth grade (Hill and Jones 2018) but an order of magnitude greater than the effect of having a same-race teacher (Egalite, Kisida, and Winters 2015).

Finally, in panel D, we allow the effect to vary by whether the student has the same math and reading teacher. Recall that this is intended to proxy for the student being in a self-contained class with a single teacher during the school day and, therefore, a more intense student-teacher relationship. In our fully specified models (cols. 3–6), we observe an additional test score gain from more positive teacher expectations for students with the same math and reading teacher, although it is less precisely estimated than the main impact.

We also note that the conclusion we draw from our analysis is not sensitive to the way that we allow teachers' expectations to enter our specification. In additional analyses, rather than allowing the teachers' reported expectation measure to enter as a continuous variable, we construct a simpler "high expectations" dummy equal to one if a teacher's expectation of a student is at least level II, at least level III, or equal to level IV. In three separate specifications, we then instrument for these high-expectations dummies using the same teacher optimism measure used in the main analysis. Table A3 reports the results of this exercise.[14] As in our main analysis, we observe a clear positive and statistically significant impact of higher expectations on student achievement. The coefficients on the various high-expectations dummies are larger than the coefficient on expectations

[13] Table A2 aims to profile which students are impacted by our treatment given that our instrumental variable strategy identifies a local effect. It reports covariate means for the whole sample, compliers, never takers, and always takers using the methodology of Dahl, Kostøl, and Mogstad (2014) and Marbach and Hangartner (2020). The approach requires a binary instrument, so results are reported for three alternative binary versions of the teacher optimism instrument: optimism above the 75th percentile, optimism above the 50th percentile, and optimism above the 25th percentile. Regardless of the instrument that is used (whichever of the columns is considered), the table shows that compliers, the students for whom the effect is identified, have greater probabilities of being nonwhite (either black or other race/ethnicity), have lower previous test scores, and are more likely to be economically disadvantaged, in comparison with the overall sample of students.

[14] This approach is preferred over an ordered probit or logit model given the higher-order fixed effects in our model.

TABLE 5
FALSIFICATION TEST: EFFECT OF NEXT YEAR'S TEACHER OPTIMISM MEASURE ON CURRENT PERFORMANCE

	Ordinary Least Squares (1)	Two-Stage Least Squares: Teacher Optimism Simple Version (2)	Two-Stage Least Squares: Teacher Optimism Main Version (3)
Lead (expectations)	.070***	.093***	.008*
	(.001)	(.007)	(.005)
R^2	.897	.897	.898

Note.—The dependent variable is end-of-grade test score (z-score). The "teacher optimism main version" measure is our main optimism measure, estimated as described in the text (eq. [2]). The teacher optimism simple version measure is estimated in a similar way to that denoted in eq. (2) but with no controls other than subject-grade-year fixed effects. All specifications include a full set of controls and fixed effects (as in table 3, col. 6, panel C). Robust standard errors (clustered at the teacher level) are in parentheses.
* $p < .10$.
*** $p < .01$.

from our main analysis. This is to be expected: shifting a student from having a teacher that expects no master (level I or II) to a teacher that expects mastery (level III or IV) is a larger marginal shift in expectations than simply moving up by one level (I to II, II to III, etc.).[15]

Table 5 reports the results of a placebo test aimed at assessing the validity of our results. For each student in year t, we examine the impact that the expectations of their year $(t + 1)$ teacher has on their year t EOG performance. If the instrument generates a valid causal estimate of expectations on performance, then there should be no effect of a future teacher on current performance. However, if we have failed to eliminate all possible threats to identification, it is possible that factors correlated with high performance in the current year are also correlated with high teacher expectations (or being assigned to a teacher that is generally optimistic) in the following year. To demonstrate this possibility, in column 1 of table 5, we run a simple OLS specification, regressing current EOG score on the following year's student-specific teacher expectation and the full set of controls employed elsewhere (as noted in the table). We observe a positive correlation between this year's performance and next year's teacher expectation (despite including a rich set of student fixed effects and controls). This is not surprising: a student who is unobservably highly motivated will perform well this year and will be expected to perform well in the following year. Reverse causation is also possible: if teachers are informed about their students' past EOG performance (or even communicate informally with

[15] In addition, in table A4, we test whether our results are driven by one subject (math or reading). We find that teacher expectations have a clear positive effect in both math and reading, but the effect is larger in math. The finding that the impact is more evident in math than reading is consistent with many other studies of educational interventions (e.g., Lavy and Schlosser 2011).

each other about student abilities), then a high performance on the year t EOG test could directly cause high expectations in the following year.

In column 3, we use the same 2SLS approach employed in our main results. We instrument for a student's year $(t + 1)$ student-specific teacher expectations with our constructed teacher optimism measure for the teacher they are assigned to in year $(t + 1)$. The estimated coefficient is now essentially a precise zero, revealing that there is no causal relationship between the year $(t + 1)$ teacher expectation and the year t EOG score. This is exactly what we would expect if our methodological approach generated true causal effects.

Notably, it is not simply the fact that we switched to a 2SLS approach in column 3 that led the placebo to fail where it had detected an effect in the OLS approach of column 1. Instead, the validity of the approach also depends on using the version of our constructed teacher optimism measure that takes full advantage of the available data. This can be seen in column 2, which is similar to the specification employed in column 3, except that we use the simplest version of the constructed teacher optimism measure ("simple version," as described in the previous section, which excludes teacher, student, and peer characteristics from the construction of the teacher optimism measure). As we discussed above, that measure is clearly inferior to the main version of the measure that we use in most of analyses. Column 2 shows that employing the 2SLS approach with an inferior measure detects an effect where there should not be one.

A. *Persistence of Effects of Teacher Expectations*

Next, we test the persistence of the positive impact of high teacher expectations. Do increased expectations impact a student only while they are exposed to the teacher with expectations, or is there a lasting effect that benefits students in future grades (regardless of the level of optimism of future teachers)? To test this, we repeat our main 2SLS specification (table 4, panel C, col. 6), with one change: rather than assessing the impact of a teacher's expectations in year t on a student's grade in year t, we assess the impact of the expectation that a student's year $(t - 1)$ teacher has on their performance on year t. That is, we focus on the impact of the lagged teacher expectation for each student. Otherwise, the approach is the same: we instrument for lagged student-specific teacher expectations using the lagged value of our constructed teacher optimism measure. Column 1 of table A5 reports the result; we find no evidence that having a teacher with higher expectations in year $(t - 1)$ leads to better performance in year t.

Effects of teacher expectations in repeat student-teacher matches.—We next present a related test in response to a different question. As noted in an earlier section, we might prefer to observe a teacher's expectation closer to the beginning of an academic year in order to study how that expectation—which in turn impacts students throughout the year—impacts achievement by the end of the year. The data available to us are expectations reported at the end of an academic year. As discussed above, we feel that this is a reasonable proxy for the expectations that shape teachers' behavior (and, in turn, student achievement) throughout the year. (If anything, the expectations measure we have access to is more accurate than we would like and therefore leaves less scope for expectations to influence outcomes.) Nonetheless, for a small subset of our data, we can more directly address whether having access to an expectation reported earlier in the year would dramatically alter our conclusions. In particular, a small number of students are matched to the same teacher 2 years in a row in different grades because the teacher's teaching assignment changes (i.e., not because of grade retention); previous studies indicate that repeat matches account for roughly 3% of student-year observations. For this small group of students, we can observe how teacher j's expectation of student i at the end of year t impacts that student's achievement at the end of year $t + 1$, after having still been taught by teacher j in year $t + 1$. In other words, we can study the impact of a much earlier-reported expectation on student achievement.

Our analysis on this front is similar to the specification we ran to test for a persistent effect of teacher expectations. Again, we take test scores from year t as our outcome and include teacher expectations from year $t - 1$ (instrumented for with the teacher optimism measure of the teacher in year $t - 1$). The only change relative to that specification is that we now interact teacher expectations with an indicator variable identifying students who are in the second year of a repeat student-teacher match. The interaction between lagged teacher expectations and the repeat match indicator identifies the effect of teacher j's expectations of student i reported at the end of year $t - 1$ on student i's achievement in year t, relative to any general persistent effect of teacher expectations that occurs without repeat student-teacher matches (which, as we saw in table A5, is not present). Table 6 reports the results. As in table A5, we observe no next-year effects on students not engaged in the second year of a repeat student-teacher match. For students who are in the second year of a repeat match, the estimated effect of the teacher's expectations (reported at the end of the previous year) is roughly twice the size of the estimate from our main specification (although less precisely estimated). The fact that the coefficient from this specification is larger is consistent with the notion that an earlier-reported expectation is less accurate and therefore generates more scope for teachers' (inaccurate) expectations to influence student achievement.

In table 6, column 2, we consider whether the effect of current teacher expectations depends on whether the student is in a repeat student-teacher match; we do not observe statistically precise heterogeneity in this dimension (although the magnitudes of the additional estimates are relatively large).

TABLE 6
Interaction of Teacher Expectations Effect and Repeat Student–Teacher Match Effect: Two-Stage Least Squares

	(1)	(2)
Lag(expectations)	−.002	
	(.009)	
Lag(expectations) × repeat match	.182*	
	(.099)	
Expectations		.046***
		(.009)
Expectations × repeat match		−.058
		(.097)
Repeat match	−.125*	.034
	(.067)	(.067)
R^2	.898	.898

Note.—The dependent variable is end-of-grade test score (*z*-score). All specifications include a full set of controls and fixed effects (as in table 3, col. 6, panel C). Robust standard errors (clustered at the teacher level) are in parentheses.
* $p < .10$.
*** $p < .01$.

B. *Heterogeneity in Results*

In this subsection, we test for heterogeneity in the teacher expectations effect along several dimensions motivated by the economics of education literature and the potential policy implications of our findings.

First, we test whether teacher expectations matter more for students from traditionally disadvantaged groups (e.g., ethnic minorities). Not only do these groups experience achievement gaps, but the literature discussed in the introduction documents that teachers hold systematically lower expectations of students in these groups. If teacher expectations impact achievement (as we have shown) and teachers hold lower expectations of disadvantaged groups (as others have shown), then a policy intervention aimed at increasing the expectations of teachers of disadvantaged students may reduce achievement gaps. However, to have the maximum impact on reducing achievement gaps, a successful intervention would require disadvantaged students to benefit from higher teacher expectations to the same extent as (or more than) other students.

It is with this in mind that we investigate how our estimates vary by students' ability (measured by quintiles of lagged EOG score), socioeconomic status (as indicated by free or reduced lunch eligibility), race, and gender. Table 7 reports the results of specifications where we interact the instrumented teacher expectation with indicators describing these groups. (We also interact the instrument with the same indicators in the first stage.)[16]

[16] Note that time-invariant student characteristics are absorbed by the student fixed effects, so we can identify only the interaction effect. These classifications are based on the student's eligibility to receive free lunch or reduced-price lunch at school. The exact thresholds vary year to year and depend on the size of the student's family. To provide one example, in 2012–3 (the last year in our sample), students from a family of four were eligible for

TABLE 7
Heterogeneity in Effect by Student Ability and Demographics: Two-Stage Least Squares

	(1)	(2)	(3)	(4)
Lowest quintile lag score × expectations	.063***			
	(.010)			
Second-lowest quintile lag score × expectations	.069***			
	(.009)			
Middle quintile lag score × expectations	.080***			
	(.010)			
Second-highest quintile lag score × expectations	.077***			
	(.010)			
Highest quintile lag score × expectations	.086***			
	(.017)			
Highly economically disadvantaged × expectations		.066***		
		(.007)		
Moderately economically disadvantaged × expectations		.062***		
		(.015)		
Not economically disadvantaged × expectations		.069***		
		(.008)		
Student ethnicity:				
Black × expectations			.059***	
			(.009)	
White × expectations			.069***	
			(.008)	
Other × expectations			.075***	
			(.011)	
Male student × expectations				.059***
				(.007)
Female student × expectations				.076***
				(.007)
F-test of equality: p-value	.415	.902	.405	.021
R^2	.888	.889	.889	.889

Note.—The dependent variable is end-of-grade test score (z-score). All specifications include a full set of controls and fixed effects (as in table 3, col. 6, panel C). Robust standard errors (clustered at the teacher level) are in parentheses.
*** $p < .01$.

Column 1 tests for heterogeneity by underlying ability. We find that teacher expectations matter throughout the ability distribution but with slightly larger coefficients for students with higher prior achievement. Column 2 assesses heterogeneity by socioeconomic status. Students are classified as highly economically disadvantaged, moderately economically disadvantaged, and not economically disadvantaged.[17] There we find that

reduced-price lunch if family income was below $42,463. Students (again from a family of four) were eligible for free lunch if family income was below $29,965.

[17] Michelmore and Dynarski (2017) argue that persistent eligibility for free or reduced-price lunch is a better measure of economic disadvantage than current free or reduced-price lunch status. In results available on request, we considered alternative definitions based on whether a student had been eligible for free or reduced-price lunch for sequential years, finding a similar pattern of effects and that our results are not sensitive to this aspect of the economic disadvantage definition.

the impact of expectations does not noticeably differ across socioeconomic groups. In column 3, we test for heterogeneity by race. We find that high teacher expectations have slightly larger impacts on white students than black students, although the difference between these coefficients is not statistically significant. Finally, in column 4, we consider impact differences by student gender. Female students experience greater gains than male students from more positive teacher expectations, and this difference is statistically significant (although relatively small). One interpretation of this finding is that female students may be less academically sure of themselves and more likely to internalize the assessments of those around them. Taking the results from table 7 as a whole, other than a small difference by gender, the effect of teacher expectations does not substantially differ across different groups of students.

We next test how our results relate to the "teacher like me" literature, which finds that student-teacher demographic matches (e.g., black students assigned to black teachers) generate higher student achievement. These results are presented with the caveat that the vast majority of teachers in our sample are white females, so reliably separating gender and race effects from gender-match and race-match effects can be more challenging. Dee (2004) finds that random assignment to an own-race teacher increases test scores for both black and white students, while Fairlie, Hoffmann, and Oreopoulos (2014) find similar results in the context of higher education. In table 8, we replicate this result in our data. Column 1 reports the result of a specification that includes an indicator for a student-teacher race match but excludes controls for teacher expectations. Consistent with the existing

TABLE 8
Impact of Teacher Expectations by Student-Teacher Race and Gender Match

	(1)	(2)	(3)	(4)	(5)	(6)
Specification	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Student-teacher race match	.007***	.006***	−.009			
	(.002)	(.002)	(.026)			
Student-teacher gender match				.001	.001	−.009**
				(.001)	(.001)	(.004)
Expectations		.067***	.064***		.067***	.058***
		(.006)	(.008)		(.006)	(.007)
Student-teacher race match × expectations			.005			
			(.009)			
Student-teacher gender match × expectations						.020***
						(.007)
R^2	.888	.889	.889	.888	.889	.889

Note.—The dependent variable is end-of-grade test score (z-score). All specifications include a full set of controls and fixed effects (as in table 3, col. 6, panel C). Robust standard errors (clustered at the teacher level) are in parentheses. OLS = ordinary least squares; 2SLS = two-stage least squares.
** $p < .05$.
*** $p < .01$.

literature, we find that race matches generate small positive achievement gains (0.007 standard deviations). However, the literature has not settled on an explanation of the mechanism driving this phenomenon. One proposed mechanism is clearly related to our paper: student-teacher race matches may lead to better perceptions or higher expectations of students, which may in turn lead to higher achievement (if self-fulfilling prophecies occur). Along these lines, Dee (2005) provides evidence from survey data that teachers are less likely to report that own-race students are disruptive or inattentive.[18]

We probe the possibility that expectations are a channel through which race match effects occur in table 8. These results are presented with the caveat that student-teacher race match has been shown to directly affect teacher expectations (Gershenson, Holt, and Papageorge 2016). Although our identification strategy relies on variation in underlying teacher optimism rather than identifying effects off individual differences in student-specific expectations, these estimates should therefore be interpreted cautiously. In column 2, we report the results of a 2SLS specification similar to our main specification (table 4, panel C, col. 6), except that we have added an indicator for student-teacher race matches as a control. The resulting estimates suggest that the teacher expectations effect does not by itself explain the student-teacher race match effect; the coefficient on the race match indicator is essentially unchanged across columns 1 and 2 despite controlling for (instrumented) teacher expectations in column 2. In column 3, we allow for an interaction between expectations and student-teacher race match. In doing so, we recenter the expectations measure, subtracting 2.5 from the measure so that it ranges from -1.5 to 1.5. We do so to ensure that the main effect of student-teacher race match is not identified from students with a 0 for teacher expectations (which is outside of the range of the measure) and is instead identified based on a student in between expectation levels II and III. The negative coefficient on the race-match indicator and the positive coefficient on the interaction with teacher expectations suggests an interactive effect between teacher expectations and student-teacher race matches. Student-teacher race matches only have a significant positive effect on student achievement when a teacher has high expectations of the student (level III or IV).

Student-teacher gender congruence may also affect the impact of teacher expectations. This is explored in columns 4–6 of table 8, in an analogous approach to how race matches are considered in columns 1–3. We do not observe an overall gender match effect in column 4, but the estimates in column 6 reveal that gender matches may amplify the effect of teacher expectations.

Finally, the effect of teacher expectations may vary by grade level. Partly motivated by the evidence of higher returns from earlier investments in

[18] Dee (2007) provides similar evidence on student-teacher gender matches.

child education (Currie 2001) and partly by the important role attributed to noncognitive skills (Heckman, Pinto, and Savelyev 2013) that are potentially affected by teacher expectations, we explore whether high teacher expectations are especially important in early grades when students are still developing a sense of their own capabilities. We interact our instrumented variable with the grade of the student (and interact the instrument with grade in the first stage) to test this. Results are reported in column 1 of table A6. In general, teacher expectations appear to have a larger impact in earlier grades, especially fourth and fifth grade. (Note that third grade is omitted as it is the first year of EOG testing, so there is no lagged EOG control available for third-grade students.) Column 2 of table A6 provides a more parametric approach to assessing the same question. We allow for heterogeneity in the treatment effect to change across grades linearly by interacting the instrumented variable with a linear grade trend (and interacting the instrument with a linear grade trend in the first stage). This specification also reveals a larger early-grade effect of teacher expectations, which shrinks as a student progresses through the grades.

As discussed in section III, however, grade effects may be conflated with self-contained class effects; a teacher who teaches a student for all of their subjects is likely to have a more developed relationship with the student, and therefore the impact of their expectations on the student's performance may be larger. In columns 3 and 4 of table A6, we therefore include additional interactions with a proxy for self-contained classes: having the same math and reading teacher in a given grade. The differences between columns 1 and 3 are small; the general pattern of greater effects in early grades is preserved, although it is weaker and the teacher expectations effect in eighth grade is strikingly large. Furthermore, in column 4, the interaction between teacher expectations and the grade trend is no longer negative, although the final row reveals a negative grade trend in the impact for students with the same math and reading teacher. Overall, we interpret the evidence in table A6 as weakly suggestive of a declining impact of teacher expectations as a student progresses through elementary and middle schools.

V. Conclusion

Using student-teacher-linked panel data from North Carolina, we show in this paper that a teacher who expects a student to display a higher level of mastery of the relevant subject material (e.g., mastery vs. inconsistent mastery or "superior mastery vs. mastery) causes this student to score 0.07 standard deviations higher on EOG standardized tests. Teachers' expectations are self-fulfilling. Given the endogeneity of teacher expectations in a standard education production model—there are clearly many unobservable factors affecting both student achievement and teacher expectations—our claim is based on constructing a measure of underlying teacher optimism to use as an instrument for student-specific teacher expectations. In addition, the panel nature of the data allows us to include student fixed effects in the estimating equation, meaning that identifying variation comes from changes in teacher optimism as the same student moves from one grade to the next.

Other than bringing fresh administrative data and empirical techniques to a decades-old debate, we find new evidence of heterogeneity in the effects of teacher expectations on student achievement. First, the impacts of teacher expectations are generally similar for higher-achieving, higher-income and white students in comparison with lower-achieving, lower-income, and minority students. This means that uniform increases in teacher expectations may not shrink achievement gaps but will, at least, not exacerbate them. Second, we find that student-teacher race matches amplify the effect of teacher expectations and that student-teacher race matches generate no gains when teachers have low expectations. This provides suggestive evidence that students are more responsive to positive inputs provided by same-race teachers. And, third, the impacts of teacher expectations are generally shown to be more important in early grades. One interpretation of this finding is that teacher expectations are especially important when students are young and still developing a sense of their own capabilities.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *J. Labor Econ*. 25 (1): 95–135.

Brophy, Jere E. 1983. "Research on the Self-Fulfilling Prophecy and Teacher Expectations." *J. Educ. Psychology* 75 (5): 631–61.

Burgess, Simon, and Ellen Greaves. 2013. "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities." *J. Labor Econ*. 31 (3): 535–76.

Burke, Mary A., and Tim R. Sass. 2013. "Classroom Peer Effects and Student Achievement." *J. Labor Econ*. 31 (1): 51–82.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *Q.J.E*. 134 (3): 1163–224.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers. II: Teacher Value-Added and Student Outcomes in Adulthood." *A.E.R.* 104 (9): 2633–79.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Econ. Educ. Rev*. 26 (6): 673–82.

———. 2010. "Teacher Credentials and Student Achievement in High School a Cross-Subject Analysis with Student Fixed Effects." *J. Human Resources* 45 (3): 655–81.

Condron, Dennis J. 2007. "Stratification and Educational Sorting: Explaining Ascriptive Inequalities in Early Childhood Reading Group Placement." *Soc. Problems* 54 (17): 139–60.

Currie, Janet. 2001. "Early Childhood Education Programs." *J. Econ. Perspectives* 15 (2): 213–38.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Q.J.E.* 129 (4): 1711–52.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Rev. Econ. and Statis.* 86 (1): 195–210.

———. 2005. "A Teacher like Me: Does Race, Ethnicity, or Gender Matter?" *A.E.R.* 95 (2): 158–65.

———. 2007. "Teachers and the Gender Gaps in Student Achievement." *J. Human Resources* 42 (3): 528–54.

Dobbie, Will, Jacob Goldin, and Crystal Yang. 2016. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." Working Paper no. 22511, NBER, Cambridge, MA.

Egalite, Anna J., Brian Kisida, and Marcus A. Winters. 2015. "Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement." *Econ. Educ. Rev.* 45:44–52.

Evans, Judith T., and Robert Rosenthal. 1969. "Interpersonal Self-Fulfilling Prophecies: Further Extrapolations from the Laboratory to the Classroom." *Proc. Ann. Convention American Psychological Assoc.* 4 (1): 371–72.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor like Me: Race and Ethnicity Interactions in the Classroom." *A.E.R.* 104 (8): 2567–91.

Ferguson, Ronald F. 2003. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." *Urban Educ.* 38 (4): 460–507.

Figlio, David N. 2005. "Names, Expectations and the Black-White Test Score Gap." Working Paper no. 11195, NBER, Cambridge, MA.

Francis, Dania V., Angela CM de Oliveira, and Carey Dimmitt. 2019. "Do School Counselors Exhibit Bias in Recommending Students for Advanced Coursework?" *BE J. Econ. Analysis and Policy* 19 (4): 1–17.

Gentrup, Sarah, Georg Lorenz, Cornelia Kristen, and Irena Kogan. 2020. "Self-Fulfilling Prophecies in the Classroom: Teacher Expectations, Teacher Feedback and Student Achievement." *Learning and Instruction* 66:101296.

Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2016. "Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations." *Econ. Educ. Rev.* 52:209–24.

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Q.J.E.* 132 (3): 1219–60.

Goldhaber, Dan. 2007. "Everyone's Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness?" *J. Human Resources* 42 (4): 765–94.

Hanna, Rema N., and Leigh L. Linden. 2012. "Discrimination in Grading." *American Econ. J. Econ. Policy* 4 (4): 146–68.

Harber, K. D., J. L. Gorman, F. P. Gengaro, S. Butisingh, W. Tsang, and R. Ouellette. 2012. "Students' Race and Teachers' Social Support Affect the Positive Feedback Bias in Public Schools." *J. Educ. Psychology* 104 (4): 1149–61.

Harris, Douglas N., and Tim R. Sass. 2009. "The Effects of NBPTS-Certified Teachers on Student Achievement." *J. Policy Analysis and Management J. Assoc. Public Policy Analysis and Management* 28 (1): 55–80.

Harris, Monica J., and Robert Rosenthal. 1985. "Mediation of Interpersonal Expectancy Effects: 31 Meta-Analyses." *Psychological Bulletin* 97 (3): 363–86.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *A.E.R.* 103 (6): 2052–86.

Hedges, L. V., and A. Nowell. 1998. "Black-White Test Score Convergence since 1965." In *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips, 149–81. Washington, DC: Brookings Institution.

Hill, Andrew J. 2017. “The Positive Influence of Female College Students on Their Male Peers.” *Labour Econ*. 44:151–60.
Hill, Andrew J., and Daniel B. Jones. 2018. “A Teacher Who Knows Me: The Academic Benefits of Repeat Student-Teacher Matches.” *Econ. Educ. Rev*. 64:1–12.
———. 2020. “The Impacts of Performance Pay on Teacher Effectiveness and Retention: Does Teacher Gender Matter?” *J. Human Resources* 55 (1): 349–85.
Jussim, Lee. 1989. “Teacher Expectations: Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy.” *J. Personality and Social Psychology* 57.3 (1989): 469.
Jussim, Lee, and Kent D. Harber. 2005. “Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies.” *Personality and Soc. Psychology Rev*. 9 (2): 131–55.
Jussim, Lee, Stacy L. Robustelli, and Thomas R. Cain. 2009. “Teacher Expectations and Self-Fulfilling Prophecies.” In *Handbook of Motivation at School*, edited by Kathryn R. Wentzel and Allan Wigfield, 349–80. New York: Routledge.
Kinsler, Josh, and Ronni Pavan. 2016. “Parental Beliefs and Investment in Children: The Distortionary Impact of Schools.” Working paper, Human Capital and Econ. Opportunity Working Group, Univ. Chicago.
Kling, Jeffrey R. 2006. “Incarceration Length, Employment, and Earnings.” *A.E.R.* 96 (3): 863–76.
Lavy, Victor. 2008. “Do Gender Stereotypes Reduce Girls’ or Boys’ Human Capital Outcomes? Evidence from a Natural Experiment.” *J.P.E.* 92 (10–11): 2083–105.
Lavy, Victor, and Analia Schlosser. 2011. “Mechanisms and Impacts of Gender Peer Effects at School.” *American Econ. J. Appl. Econ*. 3 (2): 1–33.
Link, Charles R., and Edward C. Ratledge. 1979. “Student Perceptions, IQ and Achievement.” *J. Human Resources* 14 (1): 98–111.
Marbach, Moritz, and Dominik Hangartner. 2020. “Profiling Compliers and Noncompliers for Instrumental-Variable Analysis.” *Polit. Analysis* 28 (3): 435–44.
Mechtenberg, Lydia. 2009. “Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages.” *Rev. Econ. Studies* 76 (4): 1431–59.
Michelmore, Katherine, and Susan Dynarski. 2017. “The Gap within the Gap: Using Longitudinal Data to Understand Income Differences in Educational Outcomes.” *AERA Open* 3 (1): 1–18.
Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang. 2020. “Teacher Expectations Matter.” *Rev. Econ. and Statis*. 102 (2): 234–51.
Qureshi, Javaeria A., and Ben Ost. 2018. “The Role of Families in Student Sorting to Teachers.” *J. Human Resources* 55 (2): 470–503.
Rangel, Marcos, and Ying Shi. 2020. “Teacher Biases: Evidence from Early Career Experiences.” Working paper, Syracuse Univ.
Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. “Teachers, Schools, and Academic Achievement.” *Econometrica* 73 (2): 417–58.
Rockoff, Jonah E. 2004. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *A.E.R.* 94 (2): 247–52.
Rosenthal, Robert. 1973. “The Mediation of Pygmalion Effects: A Four Factor ‘Theory.’” *Papua New Guinea J. Educ.* 9:1–12.
Rosenthal, Robert, and Lenore Jacobson. 1968. “Pygmalion in the Classroom.” *Urban Rev*. 3 (1): 16–20.
Terrier, Camille. 2016. “Boys Lag Behind: How Teachers’ Gender Biases Affect Student Achievement.” IZA Discussion Paper no. 10343, Inst. Labor Econ., Bonn.
Trouilloud, David O., Philippe G. Sarrrazin, Thomas J. Martinek, and Emma Guillet. 2002. “The Influence of Teacher Expectations on Student Achievement in Physical Education Classes: Pygmalion Revisited.” *European J. Soc. Psychology* 32 (5): 591–607.

Van den Bergh, L., E. Denessen, L. Hornstra, M. Voeten, and R. W. Holland. 2010. "The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap." *American Educ. Res. J*. 47 (2): 497–527.
Van Ewijk, Reyn. 2011. "Same Work, Lower Grade? Student Ethnicity and Teachers' Subjective Assessments." *Econ. Educ. Rev*. 30 (5): 1045–58.
Wineburg, Samuel S. 1987. "The Self-Fulfillment of the Self-Fulfilling Prophecy." *Educ. Res*. 16 (9): 28–37.
Zavodny, Madeline. 2013. "Does Weight Affect Children's Test Scores and Teacher Assessments Differently?" *Econ. Educ. Rev*. 34:135–45.